10 STATISTICAL CALCULATIONS

NOTE: Before a statistical calculation is applied the DNA profile will be evaluated to determine if the individual of interest is included or eliminated as a possible source of the genetic material. Refer to Chapter 9, Interpretation Of Powerplex® 16 BIO System PCR Amplification Results, for guidelines on interpreting DNA profiles. Only if the individual of interest cannot be eliminated as a possible source of the genetic material will a statistical calculation be performed. The statistical approach that is applied will depend on the circumstances of the case and the criterion addressed in the remainder of this chapter.

10.1 The frequency of occurrence of allele fragments reported as being consistent is determined for each polymorphic locus within a racial group. Using the allele designations specified in Chapter 9, Interpretation Of Powerplex® 16 BIO System PCR Amplification Results, determine the frequency for each allele from the appropriate population database.

10.2 The frequency associated with a particular pattern of alleles from a sample is based upon principles of Hardy-Weinberg equilibrium.

10.2.1 If a single source sample under analysis demonstrates two alleles, the genotypic frequency at a particular locus is determined by the equation 2pq, where p and q represent the frequencies of allele #1 and #2[1].

10.2.2 If a single source sample under analysis consists of a single allele, the genotypic frequency at a particular locus is determined by the equation $p^2 + p(1-p)\theta$, where $\theta = 0.01$ and p represents the frequency of the allele [1].

10.2.3 If a known sample consists of more than two alleles at a particular locus, no frequency data will be generated for that locus.

10.3 A random match probability will be determined for single source samples using the product rule. The product rule may also be applied for a mixture sample when the complete major contributor profile can be determined by subtracting the contribution of the known donor from the mixture. Exception: If the complete major contributor profile can be determined at all but one locus, where there is an overlapping pattern between the victim and suspect, a random match probability can be applied by considering the locus inconclusive for statistical purposes. Otherwise a "Likelihood Ratio or a "Combined Probability of Inclusion" discussed below must be applied.

10.4 Likelihood ratios will be used to calculate the match probability for the foreign DNA profile from an intimate sample (i.e., a biological sample that is known to have originated from one of the individuals involved, such as a vaginal swab collected from the victim) when a mixture of DNA profiles from TWO individuals is observed and the entire foreign DNA profile cannot be determined by subtracting the contribution of the known donor from the mixture profile. The "Likelihood Ratio" frequency (L) is determined by the equation [1, 2, 9, 10]:

$$L = \frac{P(E/C_X)}{P(E/C_Y)}$$

L compares two hypotheses or scenarios – informally they are 1) what is the probability of the DNA evidence if the prosecution proposition is true? and 2) what is the probability of the DNA evidence if the defense proposition is true? L provides a measure of how many times more characteristic of (1) the genetic evidence than of (2).

Where: C_X is the first explanation (the alleles of the mixed DNA profile attributed to unknown contributors under explanation C_X), where X is the number of unknown contributors under explanation C_X.

C_Y is the second explanation (the alleles of the mixed DNA profile attributed to unknown contributors under explanation C_y), where Y is the number of unknown contributors under explanation C_y.

$P(E/C_X)$ is the probability of the DNA profile (E) using explanation C_X

$P(E/C_Y)$ is the probability of the DNA profile (E) using explanation C_Y

Refer to Appendix I for specific formulas used for calculating likelihood ratio frequencies.

NOTE: Because the likelihood ratio calculation takes into account all possible contributors involved in the mixture and therefore provides a conservative probability, θ is not used in the calculation.

10.4.1 When the victim and suspect share the same two alleles (i.e., 12,13), if the victim's contribution to the mixture is the minor component throughout the mixture profile and the type that is shared by the victim and suspect is consistent with the major contributor (based upon the intensities of the alleles in relationship to the rest of the DNA profile) the alleles will be considered to have originated from the foreign contributor and will be used in the calculation.

Example: Victim - 12,13
Suspect - 12,13
Evidence - 12,13 [both alleles will be considered as unknowns]

10.4.2 If four alleles are observed in a mixture, the alleles that are foreign to the victim will be used in the calculation.

Example: Victim - 5,8
Suspect - 7,9
Evidence - 5,8,(7),(9) [the 7 and 9 alleles should be considered as unknowns]

* Alleles in parentheses ( ) are lesser in intensity than the other alleles.

10.4.3 Once the profile has been evaluated and it has been determined that the individual of interest cannot be eliminated, if any of the alleles for the individual of interest is missing at a particular locus, the locus will not be used in the calculation. However, the results at that locus may be used in the overall decision of an inclusion or exclusion.

Example: Victim - 10,11
Suspect - 6,8
Evidence - 10,11,(6) [this locus will not be used in the overall calculation of the likelihood ratio frequency]

10.5 Combined probability of inclusion (CPI) will be used to calculate the match probability when a complex mixture of DNA profiles from MORE THAN TWO individuals is obtained and no major contributor can be determined or a mixture of two individuals is obtained from a NON-INTIMATE SAMPLE (i.e., biological samples where the source of origin is uncertain, such as a knife containing blood found in a wooded area). The following combined probability of inclusion calculation will be used to establish the frequency of individuals who would be included as possible contributors to the mixture.

Probability of exclusion (PE) = Probability representing all genotypes not contributing to the mixture.

$PE = Q^2 + 2Q(1-Q)$ Where, Q = (1-P), frequency of alleles associated with genotypes at each locus not contributing to the mixture and P represents the sum of alleles in the mixture.

$CPE = 1 - [(1-PE_1)(1-PE_2)....(1-PE_n)]$

The combined probability of inclusion (CPI) can then be calculated by the equation:

CPI = 1 – CPE

NOTE: All individuals of interest must be present in the profile at a particular locus in order to use that locus for CPE calculations (e.g., a three person mixture consisting of the victim and two suspects, all three individuals must be present).

10.6 If it is suspected that a relative of the suspect may have left the genetic material at the crime scene, the DNA profile from the relative should be obtained whenever feasible. However if a suspected relative cannot be profiled, the following formulas should be used to determine the conditional probability that the relative has a particular genotype consistent with that of the suspect [1]:

Genotype of the suspect	Probability of the same genotype in a relative
Homozygote: A_iA_i	$p_i^2 + 4p_i(1-p_i)F$
Heterozygote: A_iA_j	$2p_ip_j + 2(p_i + p_j - 4p_ip_j)F$

F represents the kinship coefficient and p_i and p_j represent the frequency of the alleles for the race of the relative in question.

For parents and offspring, F = 1/4; for half-siblings F= 1/8; for uncle or nephew F = 1/8; for first cousins F= 1/16.

For full siblings the following formulas will be used:

Homozygote: A_iA_i	$(1 + 2p_i + p_i^2)/4$
Heterozygote: A_iA_j	$(1 + p_i + p_j + 2p_ip_j)/4$

10.7 The following are the formulas used when calculating Paternity/Relationship type calculations:

10.7.1 Random Man Not Excluded (RMNE): the frequency with which men selected at random from the same racial background as the Alleged Father would not be excluded as the Biological Father in a give testing in the Mother-Child Combination.

Mother	P	R
Child	P	**Q**
Alleged Father	Q	N

i. $RMNE = 2q - q^2$ (Where q equals the obligatory allele frequency)

Mother	P	Q
Child	P	Q
Alleged Father	Q	N

ii. $RMNE = 2(p + q) - (p + q)^2$ (Where p and q equal the obligate allele frequency)

10.7.2 Paternity Index (PI): This is the ratio of the chance that the mother and a man of the Alleged Father's phenotype produced the child (passed the obligate gene) compared to the chance that the mother and a random man produced the child (passed the obligate gene).

H_0 = Alleged father is the biological father
H_1 = Alleged father is not the biological father

When the numerator and denominator are divided by $P(R/H_0)$

$$P(H_0/R) = P(R/H_0) \ / \ P(R/H_0) + P(R/H_1)$$

Let $P(R/H_0) = X$
$P(R/H_1) = Y$

$$P(H_0/R) = 1 \ / \ 1 + Y/X$$

PI = X/Y or 1 / Y/X, also known as a likelihood ratio $(L)^{2,4}$

PI compares two hypotheses or scenarios – informally they are 1.) paternity, and 2) non-paternity. PI provides a measure of how many times more characteristic of (1) the genetic evidence is than of (2).

Example:
Mother (M) = 10, 11
Child (C) = 10, 11
Alleged Father (AF) = 10, 12

$$\text{Paternity Index} = \frac{(M_{10})(AF_{11}) + (M_{11})(AF_{10})}{(M_{10})(RM_{11}) + (M_{11})(RM_{10})}$$

$$= \frac{(1/2)(0) + (1/2)(AF_{10})}{(1/2)(RM_{11}) + (1/2)(RM_{10})}$$

$$= (AF_{10})/(RM_{11}) + (RM_{10})$$

Child	Mother	Alleged Father	Paternity Index
AA	AA	AA	$1/P_A$
AA	AB	AA	$1/P_A$
AB	AA	AA	$1/P_A$
AB	BB	AA	$1/P_A$
AB	BC	AA	$1/P_A$
AA	AB	AB	$1/2P_A$
AA	AA	AB	$1/2P_A$
AB	BB	AB	$1/2P_A$
AB	BB	AC	$1/2P_A$
AB	BC	AD	$1/2P_A$
AB	BC	AB	$1/2P_A$
AA	AB	AC	$1/2P_A$
AB	AA	BB	$1/P_B$
AB	AC	BB	$1/P_B$
AB	AA	AB	$1/2P_B$
AB	AA	BC	$1/2P_B$
AB	AC	BD	$1/2P_B$
AB	AA	BC	$1/2P_B$
AB	AC	AB	$1/2P_B$
AB	AB	AA	$1/P_A + P_B$
AB	AB	AB	$1/P_A + P_B$
AB	AB	BB	$1/P_A + P_B$
AB	AB	BC	$½(P_A + P_B)$
AB	AB	AC	$½(P_A + P_B)$

10.7.3 Combined Paternity Index (CPI): The combined paternity index is calculated by multiplying together each individual paternity index.

$$CPI = PI_1 \times PI_2 \times PI3\text{……....}PI_n$$

10.7.4 Probability of Paternity: The probability of paternity is expressed as a frequency (or percentage), incorporating the paternity index and a prior probability (i.e., 0.5) which compares the likelihood that the tested man may pass the required genes to the likelihood that an untested, unrelated random man of the same race may pass these genes.

$$P(H_0/R) = P(R/H_0) \ / \ P(R/H_0) + P(R/H_1)$$

When the equation is divided by $P(R/H_1)$[2,3,4]

Let $P(R/H_0) = X$ $PI = X/Y$
$P(R/H_1) = Y$

Probability of Paternity = (X/Y) / (X/Y) + 1

OR

Probability of Paternity = (CPI)(Pr) / (CPI)(Pr) + (1-Pr), with Pr = 0.5

Where: CPI = Combined Paternity Index
Pr = Prior Probability, Pr is 0.5

Refer to Appendix L for acceptable values of Probability of Paternity.

10.8 The following are the formulas used when calculating Paternity/Maternity/Relationship Type calculations from a single parent [5]:

Where:

p = frequency in the population
q = frequency in the population
AF_p = chance of passing p
AF_q = chance of passing q

10.8.1

Person		
Child	P	Q
Alleged Father	Q	R

$PI = (p)(AF_q) + (q)(AF_p) / 2pq = (p)(0.5) + (q)(0) / 2pq = 1/4q$

10.8.2

Person		
Child	P	Q
Alleged Father	Q	Q

$PI = (p)(AF_q) + (q)(AF_p) / 2pq = (p)(1) + (q)(0) / 2pq = 1/2q$

10.8.3

Person		
Child	P	Q
Alleged Father	P	Q

$PI = (p)(AF_q) + (q)(AF_p) / 2pq = (p)(0.5) + (q)(0.5) / 2pq = (p + q) / 4pq$

10.8.4

Person		
Child	Q	Q
Alleged Father	Q	R

$PI = (q)(AF_p) / q^2 = (q)(0.5) / q^2 = 1 / 2q$

10.8.5

Person		
Child	Q	Q
Alleged Father	Q	Q

$PI = (q)(AF_p) / q^2 = (q)(1) / q^2 = 1 / q$

10.9 The following are the formulas used when calculating Paternity/Maternity/Relationship Type Calculations from a single grandparent [5,6]:

Child	Grandmother	Grandfather	Paternity Index
AA	AA	Unknown	(1 + a) / 2a
AA	AB	Unknown	(1 + 2a) / 4a
AA	BC	Unknown	1 / 2
AA	BB	Unknown	1 / 2
AB	AA	Unknown	(1 + 2a) / 4a
AB	AB	Unknown	(a + b + 4ab) / (8ab)
AB	BC	Unknown	(1 + 4b) / (8b)
AB	BB	Unknown	(1 + 2b) / (4b)
AB	CC	Unknown	1 / 2
AB	CD	Unknown	1 / 2

10.10 The following are the formulas used when calculating Missing Person Calculations (where mother's and /father's genotypes are known) [5]:

Prob(EIM, F, Q) is the probability that the evidence would be observed given that the mother and the father were the parents of the evidence sample (Q).

Prob(EIM, F, U) is the probability that the evidence would be observed given that a random member of the population was the questioned sample (U).

LR is the ratio of the two probabilities = Prob(EIM, F, Q) / Prob(EIM, F, U)

$P_{A,}$ etc. is the estimated frequency of the "A" allele in the population

Mother	Question	Father	Prob(EIM, F, Q)	Prob(EIM, F, U)	LR
AA	AA	AA	$P_A^2 \times P_A^2$	$P_A^2 \times P_A^2 \times P_A^2$	$1 / P_A^2$
AA	AA	AB	$P_A^2 \times 2P_AP_B \times ½$	$P_A^2 \times 2P_AP_B \times P_A^2$	$1 / 2P_A^2$
AA	AB	BB	$P_A^2 \times P_B^2$	$P_A^2 \times P_B^2 \times 2P_AP_B$	$1 / 2P_AP_B$
AA	AB	AB	$P_A^2 \times 2P_AP_B \times ½$	$P_A^2 \times 2P_AP_B \times 2P_AP_B$	$1 / 4P_AP_B$
AA	AB	BC	$P_A^2 \times 2P_BP_C \times ½$	$P_A^2 \times 2P_BP_C \times 2P_AP_B$	$1 / 4P_AP_B$
AA	AB	BB	$2P_AP_B \times P_B^2 \times ½$	$2P_AP_B \times P_B^2 \times 2P_AP_B$	$1 / 4P_AP_B$
AB	AB	AB	$2P_AP_B \times 2P_AP_B \times (¼ + ¼)$	$2P_AP_B \times 2P_AP_B \times 2P_AP_B$	$1 / 4P_AP_B$
AB	AB	AC	$2P_AP_B \times 2P_AP_C \times ½ \times ½$	$2P_AP_B \times 2P_AP_C \times 2P_AP_B$	$1 / 8P_AP_B$
AB	AA	AA	$2P_AP_B \times P_A^2 \times ½$	$2P_AP_B \times P_A^2 \times P_A^2$	$1 / 2P_A^2$
AB	AA	AB	$2P_AP_B \times 2P_AP_B \times ½ \times ½$	$2P_AP_B \times 2P_AP_B \times P_A^2$	$1 / 4P_A^2$
AB	AA	AC	$2P_AP_B \times 2P_AP_C \times ½ \times ½$	$2P_AP_B \times 2P_AP_C \times P_A^2$	$1 / 4P_A^2$
AB	AC	CC	$2P_AP_B \times P_C^2 \times ½$	$2P_AP_B \times Pc^2 \times 2P_AP_C$	$1 / 4P_AP_C$
AB	AC	BC	$2P_AP_B \times 2P_BP_C \times ½ \times ½$	$2P_AP_B \times 2P_BP_C \times 2P_AP_C$	$1 / 8P_AP_C$
AB	AC	AC	$2P_AP_B \times 2P_AP_C \times ½ \times ½$	$2P_AP_B \times 2P_AP_C \times 2P_AP_C$	$1 / 8P_AP_C$
AB	AC	CD	$2P_AP_B \times 2P_CP_D \times ½ \times ½$	$2P_AP_B \times 2P_CP_D \times 2P_AP_C$	$1 / 8P_AP_C$

10.11 When performing a Missing Person (10.11) or Paternity Index (10.7.2) calculation on a partial DNA profile it is possible that an individual or an evidence sample may be identified that cannot be eliminated as a possible offspring/parent or have originated from the offspring/parent. However the individual's DNA profile or the evidence sample is missing an allele at a particular locus that is observed in the mother/father/child's DNA profile. The allele may be missing as a result of allelic dropout or a mutation. Therefore to account for both of these situations two probabilities will be provided in the Certificate of Analysis and will be calculated in the following manner:

10.11.1 To account for the possibility of allele dropout, the locus will not be used in the overall calculation.

10.11.2 To account for the possibility of a mutation, the formula addressed below will be used for the locus and factored into the overall calculation.

10.12 When applying the formulas for Missing Person (10.11) or Paternity Index (10.7.2) calculations, if a partial profile is obtained and the individual cannot be eliminated, however an allele is missing at a particular locus the probability for the locus will be calculated and reported two ways:

10.12.1 To account for the possibility of allele dropout, the locus will not be used in the overall calculation.

10.12.2 To account for the possibility of a mutation, the formula addressed below will be used.

10.13 The following formula is used when calculating Paternity/Maternity type Calculations involving a potential mutation [7, 8]:

$PI = M_{SR} / 2P_A$ M_{SR} refers to the specific mutation rate for changing allele S to R

$M_{SR} \cong \mu$ μ represents the average mutation rate reported by the American Association of Blood Banks (AABB)

$PI = \mu / 2P_A$ P_A, etc. is the estimated frequency of the "A" allele in the population

Average Mutation Rates Established by the AABB[11]

Locus	Paternal	Maternal
FGA	0.00312	0.00056
TPOX	0.00013	0.00005
D8S1179	0.00150	0.00023
VWA	0.00140	0.00033
Penta E	0.00131	0.00056
D18S51	0.00224	0.00064
D21S11	0.00148	0.00110
TH01	0.00008	0.00012
D3S1358	0.00128	0.00015
Penta D	0.00066	0.00064
CSF1PO	0.00144	0.00039
D16S539	0.00110	0.00026
D7S820	0.00119	0.00013
D13S317	0.00140	0.00041
D5S818	0.00115	0.00027

Note: When μ is used in the CPI

10.14 Procedure for rounding frequencies:

10.14.1 Allele and genotype frequencies will be carried out to 3 digits. If the forth digit is 4 or less the third digit will remain the same. If the fourth digit is 5 or greater then the third digit will be rounded up.

Example: 0.346**4** would be truncated to 0.346
0.346**7** would be rounded to 0.347

10.14.2 Mutation rate frequencies will be carried out to the number of digits in the actual mutation rate.

Example: 0.0002 will be carried out to 4 digits
0.0020 will be carried out to 3 digits

10.14.3 The frequency for the overall DNA pattern, termed a DNA profile, can be determined by multiplying together the genotype frequency obtained from each locus. The overall match

probability/Likelihood Ratio/CPI (combined probability of inclusion/combined paternity index) will be truncated to two significant figures.

Example: $8.169738341 \times 10^{-11} = 1$ in 12,240,294,100

Reported match probability: 1 in 12 billion

Refer to Section 11, Report Writing, for the specific wording used when reporting match probabilities/likelihood ratios/combined probability of inclusion/probability of paternity.

10.15 Procedure for calculating allele and genotype frequencies:

The following represents an example of data collected for a STR database and the procedures used to determine the allele and genotype frequencies. Refer to Appendix H for the Virginia Division of Forensic Science population databases.

Example: TH01 locus in Caucasian population (n = 209)

Allele Frequency:

Frequency of allele = Number of times the allele was observed out of all possible alleles for a particular locus/2n.

NOTE: Alleles that contain fewer than 5 events are defaulted to 5 events in order to provide a more conservative frequency.

Examples:

Allele 5 was observed 3 times out of 418 alleles. Therefore, the allele 5 will default to a total of 5 events (5/418) = 0.012

Allele 6 was observed 100 times out of 418 alleles (100/418) = 0.239

Allele 7 was observed 59 times out of 418 alleles (59/418) = 0.141

Allele 8 was observed 50 times out of 418 alleles (50/418) = 0.120

Allele 9 was observed 64 times out of 418 alleles (64/418) = 0.153

Allele 9.3 was observed 68 times out of 418 alleles (68/418) = 0.163

Allele 10 was observed 74 times out of 418 alleles (74/418) = 0.177

Allele 11 was observed 0 times out of 418 alleles. Therefore, the allele 5 will default to a total of 5 events (5/418) = 0.012

The sum of the individual allele frequencies should equal approximately 1.000. However, because events less than 5 are defaulted to 5, the total frequencies may not total to exactly 1.000:

(0.012) + (0.239) + (0.141) + (0.120) + (0.153) + (0.163) + (0.177) + (0.012) = 1.017

Expected Genotype Frequency:

Based on the assumption that the TH01 genetic locus is in Hardy-Weinberg equilibrium, the expected genotype frequencies are calculated from the allele frequencies, as in the following examples:

TH01 Genotype 7, 7:

(Frequency of the 7 allele)^2 + Frequency of the 7 allele(1-Frequency of the 7 allele)θ = $(0.141)^2$ + 0.141(1-0.141)0.01 = 0.021

OR

TH01 Genotype 7, 9.3:

2(Frequency of the 7 allele)(Frequency of the 9.3 allele) = 2(0.141)(0.340) = 0.096

REFERENCE:

1. The Evaluation of Forensic DNA Evidence, National Academy Press, Washington D.C., 1996.

2. Evett, Ian W. and Weir, Bruce S., 1998. Interpreting DNA Evidence, Chapter 7.

3. Gürtler, H. 1956. Priniciples of blood group statistical evaluation of paternity cases at the University Institute of Forensic Medicine, Copenhagen, Acta. Med. Leg. Soc., Liège 9:83-93.

4. Paternity Testing, American Association of Blood Banks, New Orleans, Louisiana, 1979

5. Fung, W., Wong, D.M., and Tsui, P. 1996. Determination of both parents using DNA profiling, Jurimetrics Journal, 36:337-342

6. Wenk, R.E., Traver, M., and Chiafari, F.A., 1996. Determination of sibship in any two person, Transfusion, 36:259-262

7. Gjertson, D.W., Appendix 13. The effect of an isolated single-locus inconsistency in the statistical evaluation of paternity, adapted from presentations given by Debra Endean and David Gjertson at the Promega Sponsored Statistical Workshop (September 1996) and to the English Speaking Working Group of the International Society of Forensic Haemogenetics (October 1996).

8. Endean, D. and Gjertson, D., Suspected DNA mutations: statistical approach.

9. Gill, P. *et al*. (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, Forensic Science International 112, 17-40.

10. Curran, J.M., *et al*. (2005) Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure, Forensic Science International 148, 47-53.

11. 2003 AABB Annual Report

♦**END**